



Comparison of simple augmentation transformations for a convolutional neural network classifying medical images

Oona Rainio¹ · Riku Klén¹

Received: 7 September 2023 / Revised: 7 September 2023 / Accepted: 2 January 2024
© The Author(s) 2024

Abstract

Simple image augmentation techniques, such as reflection, rotation, or translation, might work differently for medical images than they do for regular photographs due to the fundamental properties of medical imaging techniques and the bilateral symmetry of the human body. Here, we compare the predictions of a convolutional neural network (CNN) trained for binary classification by using either no augmentation or one of seven usual types augmentation. We have 11 different medical data sets, mostly related to lung infections or cancer, with X-rays, ultrasound (US) images, and images from positron emission tomography (PET) and magnetic resonance imaging (MRI). According to our results, the augmentation types do not produce statistically significant differences for US and PET data sets, but, for X-rays and MRI images, the best augmentation technique is adding Gaussian blur to images.

Keywords Convolutional neural network · Image data augmentation · Medical imaging

1 Introduction

A convolutional neural network (CNN) is a deep learning technique designed for processing image data. Due to the use of matrix convolutions, a CNN is able to understand the spatial relationships between the adjacent image points in the data instead of treating them as variables in a random order. Consequently, we can easily increase the amount of image data by such augmentation transformations that preserve the meaningful structures in the images but somehow modify each image so that is not a perfect copy of the original one. For instance, common types of augmentation include reflections, rotations, and cropping out borders.

Image augmentation is especially important on the medical field. Unlike regular photographs, tomography images of human patients are highly sensitive and private material protected by strict regulation, which is why there is a very limited selection of open-source data sets about them publicly available. Because of this, a researcher is often able to use only

a data set collected at their own institution and, even if several patients are imaged there, the physicians might not have enough time for the annotation of these images so that they can be given to a CNN. Furthermore, some diseases might be so rare that there are simply not so many cases diagnosed.

However, several augmentation techniques were originally designed for typical photographs and are therefore not necessary well-suited for medical images. Because of the bilateral symmetry of the human body, a reflection over the vertical axis of coronal or transaxial images is an intuitive approach for multiplying the amount of images of the brain or the head and neck area as the results are often nearly indistinguishable from the original images. While a reflection switching the top and the bottom of the image notably differs from the original image, it would also preserve the symmetry, which might be beneficial for a CNN trying to detect asymmetrically located targets such as tumors. A rotation does not preserve this symmetry, but it might be useful when the task is to find signs of disease in chest X-rays that are often visible as wispy white sections in the normally clear areas of the lungs. Furthermore, one simple augmentation transformation is to add blur to the images, which might produce very different results for an imaging method such as positron emission tomography (PET) that produces already blurry images without sharp borders between regions than it would for magnetic resonance imaging (MRI). In existing

✉ Oona Rainio
omrai@utu.fi

Riku Klén
riku.klen@utu.fi

¹ Turku PET Centre, University of Turku and Turku University Hospital, Turku, Finland

Table 1 The abbreviation used for the data set in tables of this paper, the image type, the diagnosis or the meaning of the positive and the negative instances, the total number of images, and the reference (Ref.) to the original source of the data for our 11 data sets

Data	Image type	Positive/Negative	Images	Ref
CoV X-ray	Coronal chest X-ray	COVID-19/ No lung infection	6000	[7, 16]
PNA X-ray	Coronal chest X-ray	Pneumonia/ No lung infection	3160	[14]
TB X-ray	Coronal chest X-ray	Tuberculosis/ No lung infection	1400	[17]
Shoulder X-ray	Shoulder X-ray	Shoulder joint replacement or internal fixation device/ No artificial parts	360	[21]
Wrist X-ray	Wrist X-ray	Internal fixation device/ No artificial parts	330	[21]
BC US	Breast ultrasound	Malignant breast tumor/ No tumor	100	[2]
BT MRI	Transaxial slice of an MRI image	Brain tumor/ No tumor	3000	[11]
HNC MRI	Transaxial slice of an MRI image	Head and neck cancer tumor/ No tumor	2230	[15]
HNC PET	Transaxial slice of a PET image	Head and neck cancer tumor/ No tumor	2230	[12]
PCa PET	Transaxial slice of a PET image	Prostate cancer tumor visible / not visible	1760	[3, 19]
Polar map	Polar map from dynamic myocardial PET perfusion	Ischemic/ non-ischemic based on finding obstructive coronary artery disease imaging	110	[24]

literature [4, 9, 10, 13, 22], these augmentations have typically been compared by using only one data set rather than analyzing more systemically the differences caused by the different imaging modalities.

Here, we study the impact of seven simple augmentation types on the predictions of a CNN performing binary classification. We compare the augmentation types both with each other and results obtained by using no augmentation. We use 11 different data sets, including lung X-ray images from patients with COVID-19 (CoV), pneumonia (PNA), or tuberculosis (TB), limb X-rays with artificial joints or internal fixation devices, breast cancer (BC) ultrasound (US) images, MRI images from patients with brain tumor (BT) or head and neck cancer (HNC), PET images from HNC and prostate cancer (PCa) patients, and myocardial polar maps with patients with heart disease.

2 Materials and methods

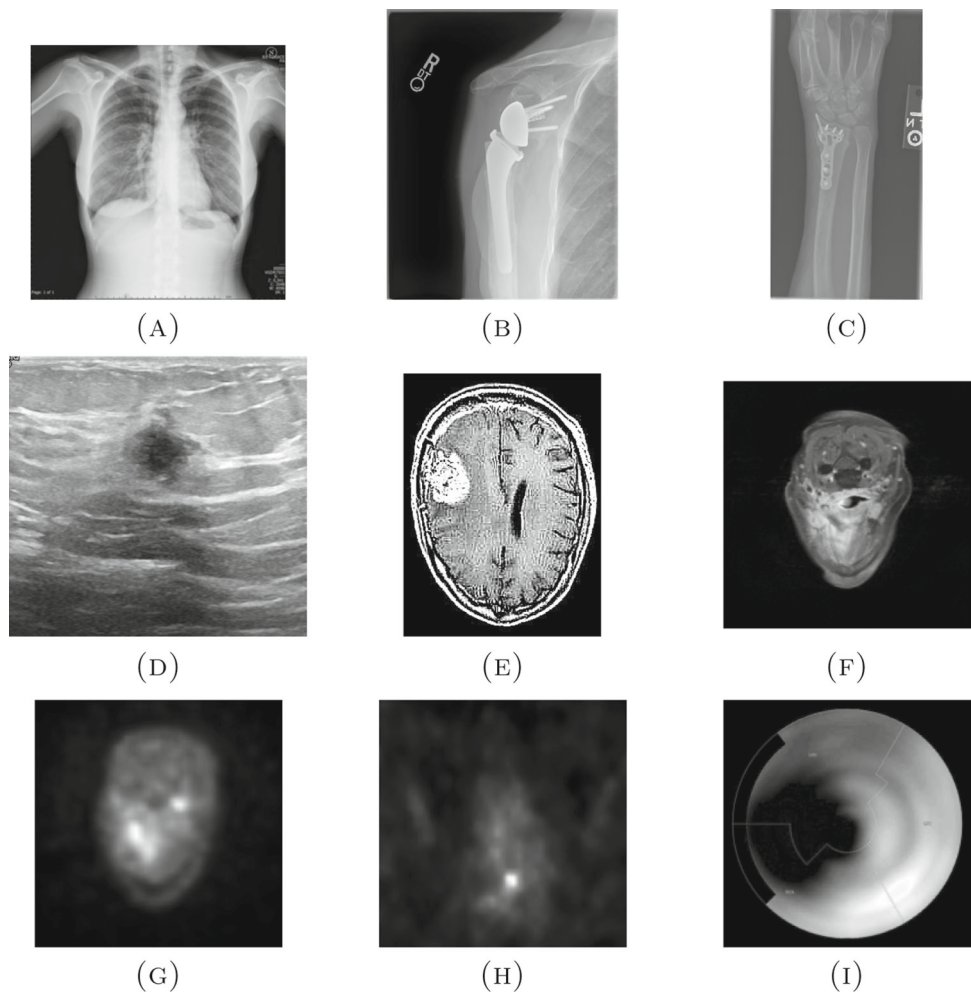
2.1 Software requirements

The CNN of this article was built and tested in Python (version: 3.9.9) [25] with the packages TensorFlow (version: 2.7.0) [1] and Keras (version: 2.7.0) [6].

2.2 Data

In this study, we use 11 different data sets, important details of which are summarized in Table 1 and the example images are shown in Fig. 1. Seven of them are created from six publicly available repositories. These repositories are COVID-19 Radiography Database [7, 16], Chest X-Ray Images (Pneumonia) [14], Tuberculosis (TB) Chest X-ray Database [17], MURA: Large Data set for Abnormality Detection in Musculoskeletal Radiographs [21], Breast Ultrasound Images Data set [2], and Br35h:: Brain Tumor Detection 2020 [11], and their links are listed in the data availability statement. They include chest X-rays of patients with either CoV, PNA, or TB, limb X-rays with different bone abnormalities, US images of BC patients, and two-dimensional (2D) MRI images of BT patients. Each data set also has similar images of negative cases. We use the MURA data set to create two smaller image sets so that one of them has shoulder X-rays with or without shoulder joint replacements or internal fixation devices, and the other one has wrist X-rays with or without internal fixation devices. Furthermore, we only include breast US images with a clearly visible tumor as positive images to our own data set so that the CNN is able to recognize them. For each data set, we choose the images so that we have equally many pos-

Fig. 1 Examples of positive images from our data sets, including a coronal chest X-ray of a COVID-19 patient **A**, a shoulder X-ray showing an artificial joint **B**, a wrist X-ray with an internal fixation device **C**, a breast ultrasound image of a patient with a malignant tumor **D**, a transaxial slice from an MRI image of a patient with a brain tumor **E**, a transaxial MRI image slice of a head and neck cancer patient **F**, the corresponding transaxial PET image slice from the same head and neck cancer patient **G**, a transaxial PET image slice showing a prostate cancer tumor **H**, and a myocardial polar map of an ischemic patient **I**



itive and negative instances and the total number of images is divisible by 10.

We also use four private data sets retrospectively collected from patients imaged at Turku PET Centre in Turku, Finland. The first two of these data sets are from 200 HNC patients, 182 of which were diagnosed with head and neck squamous cell carcinoma, while the rest of them had adenocarcinoma, adenoid cystic carcinoma, parotid cancer, or some other HNC. As a part of their treatment at Turku University Hospital, they had been referred to a PET/MRI scan in Turku PET Centre during 2014–2022, and they were imaged with either Philips Ingenuity TF PET/MRI scanner (Philips Health Care) or SIGNA PET/MRI with QuantWorks (GE Healthcare) by using ^{18}F -fluorodeoxyglucose as a tracer. The presence of cancer was confirmed for 100 patients by re-imaging or histopathological sampling and a medical doctor created three-dimensional (3D) binary tumor masks with Carimas [18] for the 100 positive PET/MRI images. We create two separate data sets so that one of them is based on MRI and the other one on PET. By using the tumor masks, we choose a total 1115 transaxial slices depicting a tumor

for both modalities and equally many random slices from the images of the 100 negative patients who did not have recurrence of the cancer. The same PET data set was also used in [12], and MRI/PET data from some of these patients have been studied in [15].

Our third private data set is from 78 PCa patients, who were imaged with Discovery MI digital PET/computed tomography (CT) system (GE Healthcare) in Turku PET Centre during 2018–2019 after a dosage of ^{18}F -prostate-specific membrane antigen-1007 (^{18}F -PSMA-1007). A physician created 3D binary masks with Carimas to denote the primary tumor in the prostate. Since there were no negative patients imaged with ^{18}F -PSMA-1007, our classification is task to detect which transaxial slices show the intraprostatic tumor and which of the slices depict healthy parts of the prostate or area outside the prostate. For this purpose, we use the binary masks to find all the slices showing the PCa tumor and choose negative slices below and above the positive slices so that we have equally many negative slices in total. We use only PET data here and fully exclude the CT data. For each transaxial PET slice, we use the same square around the prostate area

as the region of interest and crop its borders out. The data set was originally introduced in [3] and it is also studied in [19].

The last private data set is from 138 patients who had been treated at Turku University Hospital in Turku, Finland, during the years 2007-2011 and who had had stable chest pain or a similar symptom of a possible heart disease. A dynamic myocardial PET perfusion imaging was performed with Discovery VCT PET/CT scanner (GE Healthcare) by first injecting the patients with ^{15}O -labelled water as an intravenous bolus and infusing them adenosine to see how stress affected their heart. Carimas was used to combine the dynamic image sequences of each patient into one 2D image. All the patients also had an invasive coronary artery angiography on Siemens Axiom Artis coronary angiography system (Siemens), which was used to classify the polar maps as ischemic or non-ischemic based on finding obstructive coronary artery disease. Since there were 55 ischemic polar maps, we included them and equally many non-ischemic polar maps in our data set. The polar maps had been converted into RGB images with Carimas' rainbow scaling function at some point, but we converted them back into grayscale images by choosing only the value in the first channel for each pixel. More details of the original data can be found in [24].

2.3 Pre-processing and cross-validation

Our data sets had grayscale images with pixel values in [0,255], and the images were stretched into the squares of 128 pixels. The output of the CNN was therefore $128 \times 128 \times 1$ matrices whose each element was an integer from [0,255]. We created the training and test sets with fivefold cross-validation. It was done patient-wise for the all the data sets, including the ones which had multiple images from same patients. All the possible test sets contained exactly 20% of data and had equally many positive and negative images.

2.4 Convolutional neural network

In this article, we use the same CNN as in [12]. This CNN was inspired by the well-known U-Net CNN [23], which is commonly used in the medical field. The U-Net used in segmentation consists of a contracting and an expanding path, but, since our aim is binary classification instead of image segmentation, we only use the contracting path of the U-Net followed by three dense layers. This path consists of four sequences, each of which has first a convolution layer and then two maximum pooling operations. We use ReLU activation function on all the layers except the last one, which has a sigmoid function instead. We use the stochastic gradient descent with a learning rate of 0.001 as an optimizer and a validation set containing 30% of the training data. The number of epochs is 15 for shoulder X-ray, wrist X-ray, BC

US, HNC MRI, and HNC PET data sets, and 10 for the other six data sets.

2.5 Augmentation

We compare the following seven types of augmentation transformations:

- (1) Reflection over the vertical axis
- (2) Reflection over the horizontal axis
- (3) Rotation of 90 degrees in clockwise direction
- (4) Rotation of k degrees where k is a randomly chosen number from the interval $(-15, 15)$
- (5) Translation that moves the image $k_0\%$ and $k_1\%$ in the horizontal and vertical directions, respectively, where k_0, k_1 are randomly chosen from the interval $(-10, 10)$
- (6) Cropping the borders so that the size of the image decreases $k\%$ where k is randomly chosen from the interval $(0, 10)$
- (7) Adding blur from the Gaussian distribution whose standard deviation σ is randomly chosen from the interval $(0.5, 1.5)$

The augmentation transformations are tested separately so that we only use one of them during each iteration round performed. With these transformations, we create exactly one copy of every image in the training data so that we double the amount of data used for training. The first three transformations always produce the same end result from a given image, but there is some variation in the augmented images created with the other four. See Fig. 2 for an example of each augmentation type.

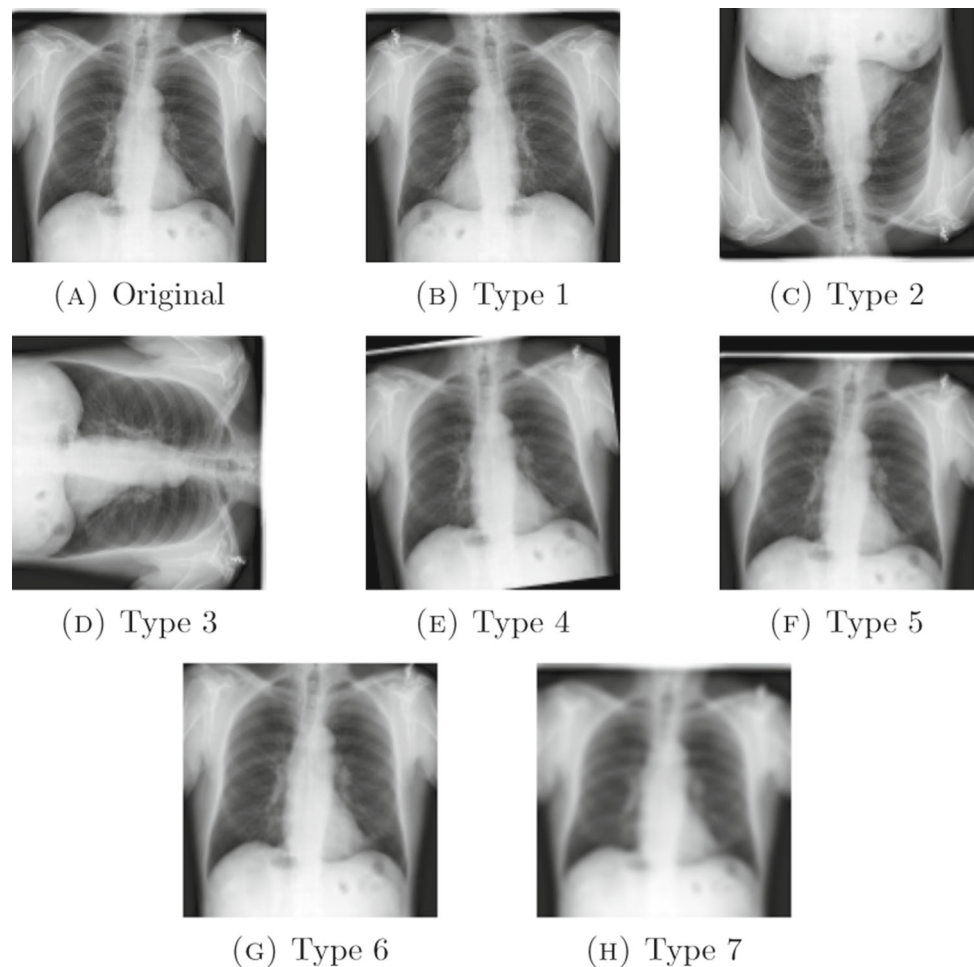
2.6 Evaluation metrics

To convert the numerical output of the CNN into binary classifications, we compute the Youden's threshold [26] from the predictions of the training data. This threshold is the one that maximizes the sum of sensitivity (percentage of positive instances classified correctly) and specificity (percentage of negative instances classified correctly). After that, we can compute the accuracy of the predictions of the test set. We also consider the receiver operating characteristic (ROC) curve, which is sensitivity plotted against the false positive rate (percentage of negative instances classified incorrectly), and compute the area under ROC curve (AUC). The AUC value can be used as a evaluation metric, but, unlike accuracy, it does not depend on the choice of the threshold.

2.7 Structure of the experiment

For each data set and augmentation type, we repeat the process of initialing the CNN, training it with augmented data,

Fig. 2 An original X-ray image of a negative patient from the CoV X-ray data set and the new versions of this image created with the seven different types of augmentation



and predicting the contents of the training and test sets for 30 times. This means six repeats of each different test set of the fivefold cross-validation. For each data set, we also run 30 iteration rounds without using any augmentation. The results are evaluated by their accuracy and AUC, and the values of these metrics are compared with Wilcoxon test with 1% level of significance so that we can estimate whether the differences between the augmentation options are statistically significant or not.

3 Results

Our results are summarized in three tables: Table 2 contains the medians of the 30 values of accuracy computed from the test set predictions given by the CNN for each data set when either no augmentation or one of the seven augmentation types presented in Subject. 2.5 is used for the training data. Table 3 has the similar medians but for AUC instead. Table 4 tells us which of the other augmentation options produce significantly higher median accuracy or AUC than augmen-

tation option in the given column for each data set according to the Wilcoxon tests with 1% level of significance.

According to Table 4, there are very few statistically significant differences between the different augmentation options for the one US and the three PET data sets. However, what augmentation is used is much more important for X-ray and MRI data sets. The use of non-augmented data leads significantly more inaccurate predictions than the seven types of augmentation for all five X-ray data sets and the BT MRI data set. The best choice seems to be augmentation by adding Gaussian blur (augmentation type 7), which gives the best accuracy or AUC values for five data sets (CoV X-ray, PNA X-ray, shoulder X-ray, BT MRI, and polar map) and has never significantly lower accuracy nor AUC than any of the augmentation types. According to our results, other good augmentation types are cropping the image (type 6), slight rotation (type 4), and translation (type 5).

Our results also reveal that the reflection over the vertical axis (type 1) only gives the highest accuracy or AUC for HNC PET and MRI data sets. It is still often better than the reflection over the horizontal axis (type 2): The first option works significantly better for CoV X-ray, TB X-ray, wrist X-

Table 2 Median accuracy (%) of the predictions of the CNN from test sets of the different data sets when either no augmentation (denoted by 0) or one of the seven different augmentation types listed in Subsect. 2.5 is used

Data	Augmentation type							
	0	1	2	3	4	5	6	7
CoV X-ray	80.8	83.0	81.2	81.6	83.1	83.0	82.6	83.7
PNA X-ray	89.6	90.1	92.0	91.2	91.4	91.8	91.8	91.6
TB X-ray	88.9	87.9	87.0	89.3	90.2	90.0	90.4	90.0
Shoulder X-ray	73.6	75.0	74.3	76.4	75.0	75.0	76.4	76.4
Wrist X-ray	75.8	78.0	72.7	77.3	78.8	78.8	79.5	78.0
BC US	75.0	75.0	75.0	75.0	77.5	75.0	75.0	75.0
BT MRI	85.3	85.5	91.3	88.2	91.5	88.9	88.7	92.4
HNC MRI	62.3	64.3	59.8	61.0	62.9	61.8	62.8	63.3
HNC PET	76.3	77.1	70.4	74.6	75.0	72.8	73.9	73.7
PCa PET	77.3	78.4	78.4	78.7	77.1	79.3	77.3	78.6
Polar map	77.3	77.3	77.3	77.3	72.7	72.7	81.8	81.8

The maximum(s) of each row are in bold, the values that are below maximum but above minimum and greater than or equal to the median of each row are both in bold and in italic, the values below the median but above the minimum are in italic, and the minimums are in regular font

Table 3 Median area under receiver operating curve values (%) of the predictions of the CNN from test sets of the different data sets when either no augmentation (denoted by 0) or one of the seven different augmentation types listed in Subsect. 2.5 is used

Data	Augmentation type							
	0	1	2	3	4	5	6	7
CoV X-ray	89.3	92.0	90.9	91.2	92.4	91.7	90.6	92.6
PNA X-ray	95.9	96.5	97.8	97.6	98.0	97.8	97.8	98.1
TB X-ray	96.1	95.8	94.7	95.9	96.6	96.8	96.5	96.4
Shoulder X-ray	79.4	83.1	81.2	83.2	81.1	82.9	83.2	83.7
Wrist X-ray	81.8	84.9	81.0	83.6	86.3	85.6	87.3	86.0
BC US	77.5	78.0	77.5	79.0	82.0	80.0	80.5	78.5
BT MRI	92.6	92.9	97.1	94.5	96.8	95.1	96.2	97.5
HNC MRI	65.3	69.7	63.6	65.9	66.3	65.9	67.4	68.2
HNC PET	85.4	85.4	80.2	84.2	85.1	82.2	81.9	82.5
PCa PET	85.5	86.0	85.7	86.2	85.1	86.0	85.2	84.9
Polar map	83.9	87.6	86.0	88.0	86.0	88.0	86.0	87.6

The fonts are as in Table 2

ray, and HNC MRI data sets, while the second is significantly better only for PNA X-ray and BT MRI data sets. However, neither of these reflections is significantly better than the slight rotation (type 4), which works better than at least one of reflection types for four X-ray data sets and the BT MRI data set.

4 Discussion

We expected that the reflection over the vertical axis would have worked very well due the bilateral symmetry of the human body, but it performed quite poorly in comparison of the random rotation of 15 degrees or less. In research by Rama et al. [22], augmentation based on reflections also led the least accuracy in classification of lung X-rays of TB patients (different data set than here). However, the reflection over the vertical axis worked the best on the HNC MRI and

PET data sets which contain highly heterogeneous images of the human head and neck area and are therefore difficult to classify correctly.

Adding the Gaussian blur was clearly the best method according to our results, and in particular, it worked very well on all the five X-ray data sets and the MRI BT data set. For instance, the AUC of predictions of the MRI BT data augmented with this method was 97.5, while the use of non-augmented data resulted in a statistically significantly lower AUC of 92.6. In earlier research by Haekal et al. [10], both Gaussian and Perlin noise worked well when classifying X-rays of lung cancer patients. However, adding Gaussian blur did not work so well on US nor PET images, possibly because the targets in these images do not have clear boundaries and can be only detected because they are lighter or darker shade than their environment, as can be seen in Subfigures (D), (G), and (B) in Fig. 1. It was also noted by Hussain et al. [13]

Table 4 The numbers in each cell refer to the augmentation options giving statistically higher median accuracy or area under receiver operating curve (AUC) than the augmentation option in the given column according to Wilcoxon tests with 1% level of significance

Data	Augmentation type							
	0	1	2	3	4	5	6	7
CoV X-ray	1-7		1,4,5,7	1,4,5,7			2	
PNA X-ray	2-7	2-7		6				
TB X-ray	4	4,5,6,7	0,1,3-7	4				
Shoulder X-ray	5-7							
Wrist X-ray	4,5-7		1,4-7					
BC US	1							
BT MRI	2-7	2-7		2,4,7		2,4,7	2,4,7	
HNC MRI			<i>1</i>	5				
HNC PET								
PCa PET								
Polar map						7		

0 means no augmentation and the numbers 1–7 are the seven augmentation types listed in Subsect. 2.5. A number in bold means that both accuracy and AUC are significantly higher according to the Wilcoxon tests, a number italicized means that only accuracy is significantly higher, and a number in regular font means that only AUC is significantly higher. The Wilcoxon test is performed for the values of accuracy and AUC computed from the predictions of the CNN for the test sets of the 11 different data sets

that adding Gaussian noise works poorly for mammography images in comparison with reflections.

Natural continuation of this study would be extending the research on 3D medical images. However, there are some challenges: There is a very limited amount of publicly available data sets with 3D medical images that also have similar negative images so that they can be used for classification. Furthermore, training CNNs for 3D data requires much more data, which might set too high requirements on the time of the tests and computational efficiency or memory of the computer.

Another question for further study would be the comparison of between these simple transformations and the more complicated augmentation types specifically designed for medical data. For instance, a general adversarial network (GAN), a type of a neural network generating synthetic samples of the original images introduced by Goodfellow et al. [8], has been applied very much deep learning tasks related to medical images [5]. While GANs have been noted to lead to better results than some affine transformations at least for certain data sets [4, 9], the best methods here might be better-suited for some imaging modalities. Also, a new method of augmentation based on the use of conformal mappings was recently introduced [20].

5 Conclusion

In this article, we compared the impact of seven simple augmentation techniques on the accuracy of the predictions of a CNN in different binary classification tasks related to medical images. We used several data sets, most which were contained either X-rays of patients with lung infections or

internal fixation devices or other imaging modalities of cancer patients. According to our results, the best method of augmentation for X-rays and MRI images is adding Gaussian blur to the images but also slight rotation of 15 degrees or less, cropping the image, and translation work quite well.

Author Contributions O.R. wrote manuscript, did the computational tests, and prepared the figures and tables while R.K. supervised the project and helped editing the text.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital). The first author was financially supported by the Magnus Ehrnrooth Foundation.

Data availability Four of the data sets are private due to ethical restrictions, but the other ones are from public repositories: COVID-19 Radiography Database [7, 16] <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>, Chest X-Ray Images (Pneumonia) [14] <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>, Tuberculosis (TB) Chest X-ray Database [17] <https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset>, MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs [21] <https://stanfordmlgroup.github.io/competitions/mura>, Breast Ultrasound Images Dataset [2] <https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset>, and Br35h: Brain Tumor Detection 2020 [11] <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>.

Code availability Available at (github.com/rklen/CNN_augmentation_comparison).

Declarations

Conflict of interest On the behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval and Informed consent All the patients of the private data sets were at least 18 years of age, consented to research use of

their data, and the research from their data was approved by Ethics Committee of the Hospital District of Southwest Finland.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Zheng, X.: TensorFlow: large-scale machine learning on heterogeneous systems (2015)
- Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. *Data Brief* **21**(28), 104863 (2019)
- Anttinen, M., Ettala, O., Malaspina, S., Jambor, I., Sandell, M., Kajander, S., Rinta-Kiikka, I., Schildt, J., Saukko, E., Rautio, P., Timonen, K.L., Matikainen, T., Noponen, T., Saunavaara, J., Löytyniemi, E., Taime, P., Kemppainen, J., Dean, P.B., Sequeiros, R.B., Aronen, H.J., Seppänen, M., Boström, P.J.: A prospective comparison of 18F-prostate-specific membrane antigen-1007 positron emission tomography computed tomography, whole-body 1.5 T magnetic resonance imaging with diffusion-weighted imaging, and single-photon emission computed tomography/computed tomography with traditional imaging in primary distant metastasis staging of prostate cancer (PROSTAGE). *Eur. Urol. Oncol.* **4**(4), 635–644 (2021)
- Bali, M., Mahara, T.: Comparison of affine and DCGAN-based data augmentation techniques for chest X-ray classification. *Procedia Comput. Sci.* **218**, 283–290 (2023)
- Chen, Y., Yang, X.-H., Wei, Z., Heidari, A.A., Zheng, N., Li, Z., Chen, H., Hu, H., Zhou, Q., Guan, Q.: Generative adversarial networks in medical image augmentation: a review. *Comput. Biol. Med.* **144**, 105382 (2022)
- Chollet, F., et al.: Keras. GitHub (2015)
- Chowdhury, M.E.H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Al-Emadi, N., Reaz, M.B.I., Islam, M.T.: Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* **8**, 132665–132676 (2020)
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27**, 2672–2680 (2014)
- Guan, S., Loew, M.: Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. *J. Med. Imaging* **6**(3), 031411 (2019)
- Haekal, M., Septiawan, R.R., Haryanto, F., Arif, I.: A comparison on the use of Perlin-noise and Gaussian noise based augmentation on X-ray classification of lung cancer patient. *J. Phys. Conf. Ser.* **1951**, 012064 (2021)
- Hamada, A.: Br35h: brain tumor detection 2020, version 12, accessed on Feb 24th (2023). <https://www.kaggle.com/ahmedhamada0/brain-tumor-detection>
- Hellström, H., Lieder, J., Rainio, O., Malaspina, S., Kemppainen, J., Klén, R.: Classification of head and neck cancer from PET images using convolutional neural networks. *Sci. Rep.* **13**, 10528 (2023)
- Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. In: *AMIA Annual Symposium Proceedings*, vol. 2017, pp. 979–984 (2018)
- Kermany, D.S., Goldbaum, M., Cai, W., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **172**(5), 1122–1131 (2018)
- Lieder, J., Hellström, H., Rainio, O., Murtojärvi, S., Malaspina, S., Hirvonen, J., Klén, R., Kemppainen, J.: Automatic segmentation of head and neck cancer from PET-MRI data using deep learning. *J. Med. Biol. Eng.* **43**(5), 532–540 (2023)
- Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Maadeed, S.A., Zughaier, S.M., Khan, M.S., Chowdhury, M.E.: Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Comput. Biol. Med.* **132**, 104319 (2021)
- Rahman, T., Khandakar, A., Kadir, M.A., Islam, K.R., Islam, K.F., Mahbub, Z.B., Ayari, M.A., Chowdhury, M.E.H.: Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. *IEEE Access* **8**, 191586–191601 (2020)
- Rainio, O., Han, C., Teuvo, J., Nesterov, S.V., Oikonen, V., Pirola, S., Laitinen, T., Tättäläinen, M., Knuuti, J., Klén, R.: Carimas: an extensive medical imaging data processing tool for research. *J. Digit. Imaging* **36**, 1885–1893 (2023)
- Rainio, O., Lahti, J., Anttinen, M., Ettala, O., Seppänen, M., Boström, P., Kemppainen, J., Klén, R.: New method of using a convolutional neural network for 2D intraprostatic tumor segmentation from PET images. *Res. Biomed. Eng.* (2023). <https://doi.org/10.1007/s42600-023-00314-7>
- Rainio, O., Nasser, M.M.S., Vuorinen, M., Klén, R.: Image augmentation with conformal mappings for a convolutional neural network. *Comput. Appl. Math.* **42**(8), 361 (2023). <https://doi.org/10.1007/s40314-023-02501-9>
- Rajpurkar, P., Irvin, J.A., Bagul, A., Yi Ding, D., Duan, T., Mehta, H., Yang, B., Zhu, K., Laird, D., Ball, R.L., Langlotz, C., Shpanskaya, K.S., Lungren, M.P., Ng, A.: MURA: large dataset for abnormality detection in musculoskeletal radiographs (2017). [arXiv:1712.06957](https://arxiv.org/abs/1712.06957)
- Rama, J., Nalini, C., Kumaravel, A.: Image pre-processing: enhance the performance of medical image classification using various data augmentation technique. *ACCENTS Trans. Image Process. Comput. Vis.* **5**(14), 7–14 (2019)
- Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Lecture Notes in Computer Science*, vol. 9351, pp. 234–241. Springer, Cham (2015)
- Teuvo, J., Schultz, J., Klén, R., Knuuti, J., Saraste, A., Ono, N., Kanaya, S.: Classification of ischemia from myocardial polar maps in 15O-H₂O cardiac perfusion imaging using a convolutional neural network. *Sci. Rep.* **12**, 2839 (2022)
- van Rossum, G., Drake, F.L.: *Python 3 Reference Manual*. CreateSpace, Amsterdam (2009)
- Youden, W.J.: Index for rating diagnostic tests. *Cancer* **3**(1), 32–35 (1950)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.